R E P O R T R E S U M E S

ED 016 520 PS 000 341

PROJECT TOBI, THE DEVELOPMENT OF A PRE-SCHOOL ACHIEVEMENT TEST. FINAL REPORT.

BY- MOSS, MARGARET H.
GEORGE WASHINGTON UNIV., WASHINGTON, D.C.

PUB DATE 31 OCT 67

EDRS PRICE MF-$0.25 HC-$1.00 23P.

DESCRIPTORS- *ACHIEVEMENT TESTS, *PRESCHOOL CHILDREN, PRESCHOOL TESTS, *STANDARDIZED TESTS, *SOCIALLY DISADVANTAGED, INSTRUMENTATION, EDUCATIONAL IMPROVEMENT, *LEARNING READINESS, EARLY EXPERIENCE, MULTIPLE CHOICE TESTS, TEST OF BASIC INFORMATION (TOBI), HEADSTART

THE TEST OF BASIC INFORMATION (TOBI) IS A 54-ITEM, MULTIPLE-CHOICE PICTURE TEST DEVELOPED TO MEASURE PREACADEMIC, SCHOOL-RELEVANT KNOWLEDGE. IT CAN BE USED TO ASSESS PROGRAM EFFECTIVENESS BY GIVING IT AS A PRE- AND POSTTEST. IT CAN BE ADMINISTERED INDIVIDUALLY, OR TO GROUPS OF UP TO 15 IF THERE IS 1 ADULT FOR 3 OR 4 CHILDREN, AND TAKES FROM 15 TO 30 MINUTES TO ADMINISTER. THE TEST ITEMS ARE URBAN-ORIENTED AND WERE SELECTED FROM A POOL OF 500. THE FINAL TEST STANDARDIZATION WAS BASED ON STAFF-ADMINISTERED TESTS OF A SAMPLE OF 539 DISADVANTAGED CHILDREN SELECTED FROM URBAN AREAS ACROSS THE COUNTRY. A KUDER-RICHARDSON 20 (KR 20) RELIABILITY OF .90 WAS OBTAINED, TOGETHER WITH A TEST-RETEST RELIABILITY OF .87. THE LAST IS QUITE UNUSUAL FOR A GROUP TEST OF 4-YEAR OLDS. THE TEST IS NOW AVAILABLE ONLY ON AN EXPERIMENTAL BASIS, BUT PUBLICATION IS BEING CONSIDERED. (DR)

ED016520

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

from OEO

Final Report: Project TOBI
The Development of a Pre-school Achievement Test

Margaret H. Moss
Department of Special Education
The George Washington University
Washington, D.C. 20006

The research reported herein was performed pursuant to a contract with the Office of Economic Opportunity, Washington, D.C.

PS000341

ACKNOWLEDGMENTS

This project could not have been carried out without the cooperation and assistance of a great many people; too many to name here. Special thanks go to those school systems and projects which permitted us to test their children, which arranged space for us, and for the cooperation of the teachers. Except in the final stages of the project, there was no data that we could provide the schools with in return for their graciousness and helpfulness.

Thanks go to members of the project staff: R.S. Cottrell, M. Lichtman, B. Williams, C. Gross and the graduate students of the Department of Special Education, who helped with the testing in the earlier stages. Mrs. Lichtman was with the project from the beginning and her services are appreciated.

I should like to thank the Head Start Office of Research and Evaluation, Office of Economic Opportunity and their research readers who saw merit in our proposal. I should like to thank Dr. Edmund W. Gordon, for having faith in our ability to construct a group test!

Lastly and most especially I should like to thank my husband, Dr. James W. Moss, for his active assistance throughout the duration of this project, who was the author of the original Test of Basic Information (TOBI), on which the present TOBI is based.

October 31, 1967

Margaret H. Moss
Director, Project TOBI

FINAL REPORT: PROJECT TOBI

The Development of a Pre-school Achievement Test

INTRODUCTION

The culturally disadvantaged child presents as yet an unmet challenge to American educators. When he enters school, he lacks those tools and skills which are necessary for him to learn and progress as he should. Basic areas of deficit are (a) language and related skills, and (b) basic information and concepts on which subsequent school learning can be built. Not all of the early experiences of the disadvantaged child are handicapping to him in school, nor are all children from similar backgrounds handicapped. Yet, the lack of specific and necessary information frequently results in educators making inappropriate generalizations which are detrimental to the best interests of the child. The educator has no specific information upon which to build a classroom technology. Without such information he is forced to proceed by trial and error methods for success that is not always forthcoming.

For many years academic achievement tests have been used to assess the development of academic skills in children as they progress through school. These tests have been useful for determining curricular changes, for assessing the effectiveness of instructional programs, and for permitting the proper placement of children within the school setting. Such tests are obviously oriented toward the kind of achievement expected of typical children enrolled in regular school classes.

With the advent of preschool education programs, and particularly the massive nation-wide Head Start program, there is now a need for a pre-academic achievement test which would serve the same purposes at this level as are served by academic achievement tests at subsequent levels. Information, with respect to grouping children at this age level, has been partly supplied in the past through the use of readiness tests. Readiness tests do not test low enough nor do they have a sufficient number of items at the lower level to constitute a reliable sample. Further, the format, the perceptual material, and the content of readiness tests are inappropriate with the disadvantaged.

By having an appropriate preschool achievement test, educators can devise special programs, adapt curricula content, methods, and materials, and evaluate their effectiveness. Without such a test there is no adequate system for program development and evaluation, and the teachers are reduced to trial and error methods in their attempts to determine the educational needs of the children. Worse than a trial and error approach is one where teachers depend upon traditional materials with little innovation. These partially adapted programs make many assumptions about the needs of Head Start children which have not been objectively substantiated.

Purpose

This project modified an existing Test of Basic Information to make it suitable for use with Head Start children. The original instrument, developed for young mentally retarded children, proved effective in differentiating culturally disadvantaged children of lower elementary grades who had received special education training and those who had not. The test represents an innovation in educational testing since it measures pre-academic knowledge, is a group test suitable for four-and five-year-old disadvantaged children, has curricular relevant content, and is standardized on disadvantaged children.

The Test of Basic Information (TOBI) has been developed to assess the amount of school-relevant information acquired by young disadvantaged children by the time they enter preschool or kindergarten programs. It's range of difficulty is sufficient to permit re-testing to measure program effects. The test helps the teacher and educator find out how much a child has learned about his neighborhood and the things around him. TOBI was developed for use with boys and girls who have a limited experiential background and would therefore also be suitable with certain handicapped children.

TOBI has been constructed to be administerable to individuals or to groups; though, at the present time, only group norms are available. It is not necessary that a psychologist administer the test, since the instructions are of such a nature as to permit the teacher to do this. When doing group testing, the size of the group depends upon a) the skills and the behavior of the children in the group, b) the availablility of extra help, and c) the physical conditions for testing. Experience suggests that there needs to be one adult for every three or four children being tested. It appears likely that the group should not exceed the number of children with which the examiner can maintain visual and auditory contact. This would probably be a group not exceeding fifteen.

TOBI is a picture test, to which the child responds by either marking the appropriate picture of a set of four or by pointing to the appropriate picture depending on whether the test is administered individually or in groups. TOBI consists of 54 test items and four demonstration items. The test is not timed and takes from 15 to 30 minutes depending upon whether it is individually or group administered.

The 54 test items of the final test book were selected as the most reliable and the best discriminating in a pool of 500 items. One consideration which influenced the selection of items for the final test book was a need to have the curricular areas of science, math, and social studies represented. In "balancing" the items in the test, efforts were also made to avoid visual and content redundancies.

PROCEDURES

Test Development

The first step in the development of the test was the development of test items. These were developed by the project staff using various resources such as a) children's books, b) books on child development, nursery school and kindergarten curricula, and c) observations of Head Start program.

In developing the TOBI, not only were 500 items developed, but eight different test book formats were used. The test books varied with respect to the number of items on the page, color of pages, specific drawings used, number of alternatives, etc. The first seven forms were designated A through G. Test forms Y, Z, ZZ, and the final TOBI test use the same book format.

Subjects

These earlier forms were tried with poverty children (i.e., children in Title I or Head Start programs) in various locales. See Table I for sex and geographic characteristics of samples used with each test form version.

TABLE I

Test Form Versions Showing Characteristics of Population Samples by Sex and Geographic Area

Test Forms	Areas	Males	Females	Total
A	Prince George's Co.	22	23	45
B	Prince George's Co.	19	11	30
C	Prince George's Co.	8	14	22
D	Fairfax and Arlington Counties	32	37	69
E	Fairfax County	*	*	33
	Arlington County	2	3	5
	Montgomery County	*	*	20

Test Forms	Areas	Males	Females	Total
F	Montgomery and Fairfax Counties	24	20	44
G	Montgomery and Fairfax Counties	16	13	29
Y	Montgomery County	6	7	13
	Baltimore	36	32	68
	Arlington County	42	35	77
	Fairfax County	15	22	37
	Tennessee	42	51	93
Z	Montgomery County	25	35	60
	Tennessee	33	49	82
ZZ	Birmingham, Ala.	42	29	71
Final	Fresno, California	35	30	65
	New York City	16	24	40
	Chicago	38	37	75
	Prince George's Co.	24	23	47
	District of Columbia	41	54	95
	Montgomery County	151	163	314
Total (all subjects)				1,434

* Where there is missing data, the information is no longer available.

The tests were administered by the project staff and by graduate students at The George Washington University. The testing done in Tennessee was arranged and carried out under the direction of Dr. James O. Miller, George Peabody College. Montgomery County staff assisted with the standardization sample testing in their county.

Test Standardization

For standardization purposes the TOBI was administered during the months of June and July, 1967. Initially, it had been the intent to sample the Head Start population on the basis of some rationale. Not only was there little concensus as to what the normative population should consist of, there were also difficulties with respect to the availability of population samples. It had been hoped that the amount of time a child had been in a program could be held constant, but this was not possible.

All testing was done by members of the staff, with groups of five to fifteen children. It had been originally planned that the teachers would test the normative population. However, when arranging with an agency or a school system to test the children in their program, it was not possible prior to arriving in that city to determine specifically which teachers would be involved. It was impossible, therefore, to forward the test materials that one needed to become familiar with the procedures prior to administering the test. Since testing had to be carried out immediately upon arrival of the staff member, it became a necessity, unfortunately, that the staff member do the testing and the teachers and their aids assist.

RESULTS

500 Items and a Test

The items were constructed using the various approaches previously indicated. Some of the items were mounted on cards and tried out with small samples to determine general ambiguities of both the pictures and the item statement. In some cases pictures were re-drawn and in other cases new ones made up; sometimes the wording of the item was altered; occassionaly, the item was abandoned when neither of the above two alterations were possible.

Some of the items were tried out in one of the various formats and an over-all P value (P_t) or percent of total passing was computed. Whenever the n was sufficiently large, the sample was divided into thirds and the P values for the upper third (P_u) and for the lower third (P_l) were computed. Those items whose P_u was higher than P_l and whose percent of total passing (P_t) was between .40 and .60, were considered superior items and were retained. Some items whose characteristics were not that good, were also retained out of necessity to balance the test in the various curricular areas previously mentioned.

These various procedures were repeated as new items were generated. Consecutive test forms retained the better, older items along with the new items. After test form G, only a few new items were generated. The best 95 items, and certain demonstration items, were divided into two test books (Y and Z) of 47 and 48 items each. Three new items were constructed for TOBI Y and two new items for TOBI Z to bring the item total to fifty for each test form.

TOBI Y and Z each consisted of two four-choice items per page[1] compiled into two booklets. Two booklets were used for economic reasons--one booklet was printed with four items per page and stapled and then cut in half to provide for two items per page. Items 1.1, 1.2, 2.1, and 14.1 were demonstration items. When both booklets were given consecutively and the children were successfully following instructions, it was suggested that 14.1 be given, but not "demonstrated."

The source of the items for TOBI Y and Z and the sample n's are shown in Table 2. The sample sizes for try-outs with forms F and G were smaller than planned, partly due to the bad weather experienced at that time.

TABLE 2

Source of Items and Their Sample n's for Those Items Comprising TOBI Y and Z.

	TOBI Y			TOBI Z		
Test form source	D	E	New	F	G	New
Number of items	27	20	3	19	29	2
P_u and P_l n's	24	19		15	10	
Total sample n's	69	58		44	29	

The P values of those items which were selected to make up TOBI Y and TOBI Z are shown in Tables 3 and 4, respectively.

TABLE 3

Percent Passing for Total Upper and Lower Groups for Items From Forms D and E which were selected to Comprise Semi-Final TOBI Y

Booklet One

Item	Upper Group P (P_u)	Lower Group P (P_l)	Total P (P_t)
1.1*-Demo.**			
1.2*-Demo.**	1.00	.74	.90
2.1 -Demo.	.63	.42	.57
2.2*	.71	.38	.55
3.1	.74	.21	.36
3.2	.92	.58	.71
4.1*	.88	.63	.74
4.2*	.88	.58	.74

[1]Our predilection to have two items to a page was fostered by fruitful discussions with Dr. James O. Miller, Project DARCEE, George Peabody College, Nashville, Tennessee.

TABLE 3 continued

Item	Upper Group P (P_u)	Lower Group P (P_l)	Total P (P_t)
5.1	.63	.33	.51
5.2*	.88	.58	.68
6.1	.58	.17	.39
6.2	New		
7.1	.63	.37	.45
7.2*	1.00	.63	.81
8.1	.79	.11	.45
8.2	.89	.37	.64
9.1	.84	.42	.67
9.2	.83	.50	.71
10,1*	.54	.08	.32
10.2	.54	.17	.33
11.1*	.67	.46	.59
11.2	.89	.21	.62
12.1	.96	.71	.85
12.2	.95	.53	.73
13.1	1.00	.58	.84
13.2	.88	.46	.71
Booklet One	$\overline{X}_{P_u}$ = .80	$\overline{X}_{P_l}$ = .43	$\overline{X}_{P_t}$ = .62

Booklet Two

Item	Upper Group P (P_u)	Lower Group P (P_l)	Total P (P_t)
14.1- Demo .**			
14.2	.95	.63	.72
15.1	.89	.53	.71
15.2	.83	.50	.71
16.1	.92	.71	.77
16.2	1.00	.50	.74
17.1	.83	.42	.64
17.2	New		
18.1	.89	.53	.71
18.2	.89	.37	.64
19.1	New		
19.2	.74	.26	.41
20.1	.75	.21	.46
20.2	.74	.26	.48
21.1*	.88	.29	.52
21.2*	.58	.26	.40

TABLE 3 (continued)

Item	Upper Group P (P_u)	Lower Group P (P_l)	Total P (P_t)
22.1	.74	.42	.53
22.2	.38	.13	.25
23.1*	.75	.54	.70
23.2	.83	.25	.49
24.1	.54	.33	.45
24.2	.96	.58	.75
25.1	.67	.38	.51
25.2*	.84	.31	.57
26.1*	.89	.16	.43
26.2	.96	.54	.75
Booklet two	$\overline{X}_{P_u}$ =.80	$\overline{X}_{P_l}$ =.40	$\overline{X}_{P_t}$ =.58
Total Average	μ_{P_u} =.80	μ_{P_l} =.41	μ_{P_t} = .60

*Changed (single picture changes, changes in order of pictures) from original source item.

**Demo. is an abreviation for "demonstration".

TABLE 4

Percent Passing for Total Upper and Lower Groups for Items from Forms F and G which were Selected to Comprise Semi-Final TOBI Z

Booklet One

Item	Upper Group P (P_u)	Lower Group P (P_l)	Total P (P_t)
1.1-Demo.**			
1.2-Demo.**	.93	.40	.70
2.1-Demo.**	1.00	.47	.80
2.2	.73	.20	.48
3.1	.80	.30	.59
3.2	.80	.30	.62
4.1	.70	.20	.48
4.2	.90	.10	.45
5.1	1.00	.60	.86
5.2	1.00	.40	.69
6.1	.67	.07	.36
6.2	1.00	.10	.55
7.1	.66	.47	.50
7.2	.80	.20	.59
8.1	.70	.50	.65
8.2	.70	.40	.65

TABLE 4 (continued)

Item	Upper Group P (P_u)	Lower Group P (P_l)	Total P (P_t)
9.1	.70	.20	.41
9.2	.90	.20	.62
10.1	.60	.27	.43
10.2	.73	.27	.55
11.1	.80	.40	.55
11.2	1.00	.47	.73
12.1	.66	.00	.27
12.2	.80	.10	.41
13.1	.53	.13	.30
13.2	.70	.30	.55
Booklet One	$\bar{X}_{P_u}$ = .79	$\bar{X}_{P_l}$ = .28	$\bar{X}_{P_t}$ = .55
	Booklet Two		
14.1-Demonstration			
14.2	.90	.70	.72
15.1	.80	.40	.69
15.2	.93	.33	.64
16.1	.80	.50	.69
16.2	.73	.53	.68
17.1	.80	.33	.59
17.2	.60	.20	.41
18.1	.80	.20	.59
18.2	.60	.20	.38
19.1	.70	.20	.55
19.2	.53	.00	.27
20.1	.80	.30	.55
20.2	.73	.40	.59
21.1	.67	.13	.39
21.2	1.00	.30	.72
22.1	.87	.60	.66
22.2	1.00	.20	.45
23.1	.70	.40	.62
23.2	.93	.27	.64
24.1	.60	.30	.48
24.2	.90	.20	.45
25.1	.70	.40	.52
25.2	New		
26.1	.70	.00	.45
26.2	New		
Booklet Two	$\bar{X}_{P_u}$ =.77	$\bar{X}_{P_l}$ =.31	$\bar{X}_{P_t}$ =.55
Total Average	μ_{P_u} =.78	μ_{P_l} =.29	μ_{P_t} =.55

TOBI Y and Z were now tried out on large samples of disadvantaged children. TOBI Y was tried out with 288 children; TOBI Z with 142 children. Analyses were made and a final test of 60 items, one booklet, two items per page, was printed. A sample of 84 children was obtained in Birmingham, Alabama, and this test was first tried out there. This experience indicated that several items were ambiguous and that a 60 item test was probably too long when group administered. For these reasons, and because many of the pictures needed to be re-drawn, another test, having a number of new pictures and six items less, was compiled. This test then was the final version.

The Birmingham sample consisted of 32 children who were currently enrolled in the Head Start program and 39 children who had been selected for the summer program starting in a few weeks, but who had had no program experiences. The average age of the program children was five years, four months, that of the non-program children was five years. The program children earned an average score of 36.7, with one child earning a below-chance score. The non-program children had an average score of 26.8 with eleven children, (35%) earning scores below chance. The difference in the means for these two groups, was significant at greater than the .01 level.

Test Standardization

The testing was done in the actual classrooms, or comparable facilities, at the various schools. When necessary, special provision was made for the children not being tested to be out of the room during actual testing time. Sometimes the children were tested in another available room. Occasionally, the class was of such a size that the children could all be tested together. In total, 636 children between the ages of three and a half and six and a half were tested. Of this number, 539 children constitute the standardization population. This loss of almost 100 children from the standardization population was due to the following: a) five tests could not be used due to lack of identifying data, b) ten tests could not be used because they lacked the necessary identifying information, c) tests were eliminated because the child was either less than four or more than six, d) other tests were eliminated because the child had previously had test practice with Form Z. The children younger than four and older than six would have been included in the normative data if there had been large enough sample n's. (It appeared, though, that TOBI is too difficult for many three year old disadvantaged children.) Normative data, then, was compiled only for four and five year olds.

The normative group was sampled from the following geographic areas: New York City, Prince George's, and Montgomery Counties, Maryland; District of Columbia; Chicago; and Fresno, California.

Rural schools were excluded from the sample as TOBI has been developed with urban children. This decision to focus only upon urban children occurred early in the project. It became apparent that the items chosen and the pictures selected to illustrate them had a definite urban vs. rural bias that limited the usefulness of certain items. To avoid this, the urban child became the focus for item development.

In total, the normative population consisted of 260 boys and 279 girls. Racially, the group included 112 Caucasions (including 17 Mexican-Americans) and 427 Negroes. These latter figures represent the following percentages: 21% Caucasions (including 3% Mexican-Americans) and 79% Negroes. Since a majority of the children in poverty programs in urban areas appear to be Negro, it is felt that these percentages are not too biased.

The normative group consists of four age groups. Table 5 shows the various characteristics of these groups.

TABLE 5

Sample Size and Score Characteristics of the Four Normative Age Groups

Age Groups by Year & Months	Sample N	TOBI Mean Score	Standard Deviation of Scores
4-0 to 4-5	48	21.7	7.7
4-6 to 4-11	178	28.2	9.3
5-0 to 5-5	227	32.3	8.3
5-6 to 5-11	86	34.2	7.9

F's were computed to assess the differences in the standard deviations of each age group. None were significant. An analysis of variance was computed to assess age group mean score differences. Significant differences were found for each consecutive age group, with one exception. There is no significant difference between the five year old group and the five and a half yearold group. Table 6 shows the means and standard deviations for the normative age groups by sex. There were no significant differences by age group for sex.

TABLE 6

Sample Size and Score Characteristics of the Normative Groups by Age and Sex

Age Groups by Years & Months	Sample N		Mean Scores		Standard Deviation of Scores	
	Boys	Girls	Boys	Girls	Boys	Girls
4-0 to 4-5	18	30	21.4	21.9	6.1[1]	8.6
4-6 to 4-11	90	88	27.0	29.4	9.5	8.9
5-0 to 5-5	110	117	32.4	32.2	8.8	7.7
5-6 to 5-11	42	44	34.5	34.1	7.3	8.4

1. This group had the smallest N.

The raw scores of the normative population were converted into percentile and standard scores for each 6 month group from four to six years of age. A T score (mean=50, s.d.=10) is the standard score used. Table 2 shows the raw and converted scores for each age group. (See Appendix A, Table 2)

Group Testability

Initially there was some concern about group testability and when some testing (with one of the earlier test forms) was to be carried out in one day in Baltimore, a pre-test was developed as a screening technique. Four relatively easy items were printed up on a single sheet. The first item was for demonstration purposes. A passing score was two correct of the remaining three items. Children who passed the pre-test were group tested a day or so later. All children were found to be group testable. However, there is no evidence regarding how many children failed this test who could have been tested in a group. In fact, it was observed later on that any child who is able to answer items at the level of difficulty encompassed by TOBI, is able to take the test in a group. Even children who performed very poorly on the test items, were observed to follow the mechanics of the test taking with some help and with a little practice.

Speaking more qualitatively, however, it is not known to what extent group testing affected the scores earned, nor for which children there might be an even greater or lesser effect. Plans are underway to evaluate these questions in the near future.

Reliability

Various reliabilities were computed. See Appendix A. A Kuder-Richardson 20 (KR 20) was computed using the test variance, number of test items and percentage of persons correctly answering each item. For this test, KR 20=.90 was obtained.

A most impressive reliability was the test-retest, r=.87. The population sample was 84 Head Start children in Montgomery County, Maryland, who were group tested by local school staff with the help of Project TOBI staff. Testing was carried out in July, with re-testing (again in groups) with the same test being three days to two weeks later. There appeared to be no difference accountable by age or period of retest delay.

DISCUSSION

Value of Test

The test is a value to the educator for assessing the child's knowledge of certain basic facts which are relevant to a school program. The test can be administered at the beginning and again at the end of a program or unit to evaluate program effectiveness.

The test is particularly well suited for educational purposes as the items relate directly to school subject matter. It's usefulness is further enhanced because it can be administered to groups.

One rather pleasant aspect of the test was that the teachers were very pleased with it and thought it met a very great need. A few of the items did meet with some criticisms regarding possible ambiguities and some of the pictures were considered less than to be desired.

Individual Administration

There is no question but that a test score is influenced by the way a test is administered. When TOBI is individually administered, the child simply responds by putting his finger on the picture of his choice. The examiner is able to pace the questions to suit the child's rate of responding and is able to observe more closely the child's behavior while taking the test.

Group Administration

TOBI was constructed, both in physical format and by the nature of the testing instructions, to be administered to groups. It has only two items per page, the pages are alternately colored, and the instructions, carefully lead the child to acquire a basic understanding of the testing terms and of the task he is to undertake before testing begins. However, it is important that the examiner follow the instructions, particularly with regard to administering the test at a fairly fast pace; otherwise, some of the children will proceed to color or mark additional pictures in the item or elsewhere on the page. It is also considered important that a child's card always cover the item not being asked.

Group testing is more efficient and economical for the teacher. The norms for converting the raw scores to standard or percentile scores are based on group administered tests. Group testing is particularly suitable for Head Start classes because of the reduced size of the class and the availability of aids.

Work to be Done

The TOBI needs normative data for individually administered tests. This will allow for greater flexibility in its use. The reliability of individually administered tests needs to be determined and compared with group administrated tests.

Some children take longer to acquire the test taking know-how than others. It is possible that a simple pre-test practice might be beneficial to these children and might subsequently result in a test score more closely in line with their knowledge. A first step here might be to evaluate the reliability of the first ten or so administered test items and compare this with the reliability of the remainder of the test. This would require that a test be assembled randomly from the present items as there is some bias in the way they were initially assembled.

As there is some hope that Head Start will be extended down to include three year olds, there is a need for suitable items at this age level. This would no doubt result in another form, as to add additional items to the present test would make it too long and yet not likely allow enough items at the younger age level to constitute a reliable sample. The four year olds to four years five months could be included profitably with this younger age group.

There is a great need to provide additional information at the present age levels. This could quite suitably be accomplished by providing a profile test score. This profile score would be based on sub-test performances in the areas of: social studies, math, science, and language arts.[1] A proposal has been submitted to the Office of Economic Opportunity requesting such a project.

It is desirable that TOBI be administered to middle-class children. Comparisons with these scores, particularly, would provide some information with respect to the nature and extent of the problem. Again, this would be especially helpful if deficits and assets in curricular areas could be assessed.

CONCLUSIONS

Test Now Available

TOBI is sufficiently developed and has norms which makes it usable for group testing. As the test has not been produced yet for commercial purposes, it is available only for limited research purposes, under the direction of the project staff. Exploratory discussions are being undertaken with the intent of having it ultimately published.

Additional Necessary Work

Additional necessary work has been discussed in the preceeding section. In addition to those foreseeable projects, there is some interest in developing norms for TOBI which would make it a suitable test for use with retarded children or other handicapped children. The teacher's needs for remedial educational planning for these groups is similar to that of any atypical groups where specific information is lacking and the usual school program is not adequate. I.Q. tests may help predict the child's expected rate of learning but they provide little information that a teacher can directly use in curriculum planning.

1 It was S. Engelmann, Institute for Research on Exceptional Children, University of Illinois, who suggested that we add the area of language arts to our other three areas, thus, encompassing all of the primary curricular areas.

Group Testability

TOBI presents some reliability (test-retest) figures that somewhat challenges concepts of test reliabilities for children under six years of age. This seems even more remarkable as it is a group test. No other test is known of that can reliably test four year olds in groups.

APPENDIX A

Table 1

Point Biserial, Phi and Biserial Correlations by Age Groups

Item No.	Point Biserial				Phi				Biserial			
	1	2	3	4*	1	2	3	4*	1	2	3	4*
1.1	22	38	34	39	13	28	15	22	30	55	59	66
1.2	46	46	31	23	38	34	21	17	62	60	46	35
2.1	29	35	38	16	13	25	25	22	37	45	49	21
2.2	-08	12	10	15	-10	15	08	16	-13	20	15	22
3.1	12	11	09	-00	33	11	01	-02	15	14	12	-00
3.2	15	40	29	54	03	35	27	44	21	51	39	71
4.1	16	17	18	-04	14	14	12	06	28	30	29	-07
4.2	60	40	22	36	50	28	23	31	77	52	33	57
5.1	21	35	25	48	18	31	20	40	31	44	31	60
5.2	00	31	29	38	-17	24	22	32	00	40	38	48
6.1	27	28	19	11	36	22	12	01	36	35	24	14
6.2	08	17	32	42	-07	10	39	35	11	22	41	53
7.1	40	51	40	38	32	30	30	14	51	72	58	64
7.2	43	46	45	38	35	32	27	36	55	61	65	59
8.1	35	52	50	44	36	33	32	41	44	68	72	65
8.2	05	11	24	50	04	08	17	34	06	15	31	65
9.1	20	54	41	43	12	47	30	39	27	69	54	56
9.2	18	46	22	41	07	31	13	31	23	64	30	55
10.1	54	42	36	25	51	29	19	08	67	55	53	38
10.2	14	30	39	26	-02	23	32	24	19	39	49	33
11.1	22	23	22	22	26	11	17	29	44	35	33	34
11.2	31	44	46	42	40	33	33	33	38	56	62	55
12.1	22	37	32	37	03	35	25	40	30	49	42	47
12.2	35	35	40	41	15	25	33	31	45	44	51	53
13.1	60	42	37	34	52	41	30	24	76	53	48	44
13.2	19	21	22	29	14	20	15	26	24	27	28	37
14.1	55	24	35	07	40	12	26	15	70	32	47	09
14.2	21	35	31	16	07	30	27	14	27	45	40	22

Note.--Decimal points are omitted.

*Age in Years and Months:

1 = 4-0 to 4-5
2 = 4-6 to 4-11
3 = 5-0 to 5-5
4 = 5-6 to 5-11

Table 1 (con't)

Item no.	Point Biserial				Phi				Biserial			
	1	2	3	4*	1	2	3	4*	1	2	3	4*
15.1	47	42	50	54	44	26	34	45	60	54	66	72
15.2	48	48	40	26	35	33	21	27	66	62	57	44
16.1	53	37	39	05	35	23	15	-00	67	51	61	08
16.2	-03	20	23	15	-09	07	21	20	-05	29	38	23
17.1	39	47	48	34	30	39	41	41	50	59	61	44
17.2	40	50	38	36	29	32	21	32	51	70	54	64
18.1	33	53	44	50	26	38	31	40	41	66	58	64
18.2	46	36	31	47	40	22	17	40	60	45	39	59
19.1	25	48	43	41	29	43	31	18	34	60	56	52
19.2	16	31	26	48	12	23	21	44	24	43	34	60
20.1	17	32	39	55	12	20	34	45	23	40	49	70
20.2	19	35	40	48	04	17	39	51	30	46	52	61
21.1	35	40	44	60	37	25	35	52	51	51	55	76
21.2	55	38	32	33	44	15	10	16	73	53	52	60
22.1	42	51	36	24	15	39	26	18	54	64	46	33
22.2	55	48	48	43	61	17	24	22	68	65	74	62
23.1	51	43	51	52	39	31	35	30	67	54	66	70
23.2	23	45	46	38	14	34	35	27	29	57	59	51
24.1	25	40	48	52	16	38	37	42	36	54	61	67
24.2	41	46	43	31	36	29	23	37	52	59	59	46
25.1	42	50	41	20	36	32	29	16	54	65	58	30
25.2	41	46	36	38	31	20	28	25	52	62	50	56
26.1	50	37	30	37	50	16	11	23	67	58	48	70
26.2	44	52	47	38	51	37	29	30	56	67	63	55
27.1	25	24	18	22	22	19	10	12	33	32	25	29
27.2	05	34	41	14	-17	27	37	27	06	43	52	17

*Age in Years and Months:

1 = 4-0 to 4-5
2 = 4-6 to 4-11
3 = 5-0 to 5-5
4 = 5-6 to 5-11

TABLE 2

TABLE OF GROUP ADMINISTERED STANDARD AND PERCENTILE SCORES

Raw Score	4-0 to 4-5* SS	%iles	4-6 to 4-11* SS	%iles	5-0 to 5-5* SS	%iles	5-6 to 5-11* SS	%iles	Raw Score
54-51									54-51
50					71	99+			50
49					70	99			49
48			71	99+	69	99			48
47			70	99	68	99			47
46			69	99	67	99	65	99+	46
45			68	99	65	98	64	98	45
44			67	99	64	96	62	92	44
43			66	97	63	95	61	86	43
42			65	96	62	91	60	84	42
41			64	95	61	88	59	77	41
40			63	91	59	82	57	73	40
39			62	89	58	79	56	67	39
38			61	88	57	76	55	66	38
37			59	83	56	72	54	63	37
36	69	99+	58	80	55	66	52	52	36
35	67	98	57	78	53	61	51	49	35
34	66	96	56	71	52	56	50	45	34
33	65	94	55	69	51	53	48	41	33
32	63	92	54	67	50	48	47	39	32
31	62	87	53	62	48	43	46	38	31
30	61	87	52	57	47	38	45	35	30
29	59	81	51	54	46	36	43	30	29
28	58	75	50	50	45	30	42	28	28
27	57	75	49	44	44	23	41	23	27
26	56	71	48	41	42	21	40	21	26
25	54	69	47	36	41	18	38	17	25
24	53	60	45	28	40	16	37	15	24
23	52	54	44	26	39	14	36	12	23
22	50	52	43	24	38	11	34	10	22
21	49	52	42	22	36	10	33	5	21
20	48	46	41	19	35	8	32	3	20
19	46	42	40	18	34	7	31	3	19
18	45	37	39	16	33	6	30	2	18
17	44	31	38	14	32	5	28	2	17
16	43	29	37	10	30	5	27	1	16
15	41	19	36	10	29	4			15
14	40	12	35	7	28	3			14
13	39	10	34	6	27	2			13
12	38	10	33	6	26	2			12
11	36	10	31	4	24	2			11
10	35	10	30	4	23	2			10
9	34	8	29	3	22	2			9
8	32	6	28	3	21	2			8
7	31	6	27	3	20	1			7
6	30	4	26	2	18	1			6
5	28	2	25	2					5
4			24	2					4
3			23	2					3
2			22	2					2
1			21	2					1
0			20	1					0

*Age by years and months.